知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum

E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【专稿】

科技资源描述模型和建立方法研究

◎顾复 刘杨圣彦 顾新建
浙江大学机械工程学院工业工程研究所 杭州 310027

摘要：[目的 / 意义] 科技创新是我国发展的关键途径，需要科技资源的共享和协同创新。科技资源共享是一个系统工程，需要建立科技资源的描述模型，在此基础上进行科技资源集成、评价和分享。[方法 / 过程] 提出科技资源描述模型的结构框架，包括：科技资源分类模型、科技资源元数据模型、科技资源本体模型、科技资源知识元模型、科技资源图谱模型等。其中，科技资源包括知识、数据、产品、人才、软件、硬件等资源。本文阐述了科技资源描述模型的特点和作用，并给出科技资源描述模型的建立方法。[结果 / 结论] 本文的主要贡献是：①通过科技资源描述模型的规范化，有助于不同类型的科技资源的集成分享；②通过科技资源的不同类型的描述模型的集成研究，形成科技资源描述模型的体系架构，为进行科技资源的全面系统描述提供整体解决方案，有助于解决科技资源共享难的问题；③提出科技资源描述模型的建立方法，其特点是利用新一代信息技术依靠大众共建模型，依靠科技资源描述过程的大数据智能分析技术，建立和优化科技资源描述模型。

关键词：科技资源；科技资源描述模型；模型建立方法；科技资源共享；知识图谱
分类号：G311

引用格式：顾复 , 刘杨圣彦 , 顾新建 . 科技资源描述模型和建立方法研究 [J/OL]. 知识管理论坛 , 2020, 5(2): 69-81[引用日期]. http://www.kmf.ac.cn/p/201/.

1 引言

习近平总书记在中国科学院第十九次院士大会、中国工程院第十四次院士大会上的讲话中指出：科技体制改革还存在一些有待解决的突出问题，主要是国家创新体系整体效能还不强，科技创新资源分散、重复、低效的问题还没有从根本上得到解决。

规范、合理、科学的科技资源描述方法是解决科技资源分散、重复、低效问题的有效方法之一。利用科技资源描述方法可从不同角度

基金项目：本文系国家重点研发计划资助项目“科技资源分享模型与开放分享理论”（项目编号：2017YFB1400302）和国家自然科学基金面上资助项目“产品模块化智能设计理论和技术研究”（项目编号：51775493）研究成果之一。
作者简介：顾复（ORCID: 0000-0003-3062-1935），讲师，博士；刘杨圣彦（ORCID: 0000-0003-0061-1517），博士研究生，通讯作者，E-mail: 1964001145@qq.com；顾新建（ORCID: 0000-0002-9869-704X），教授，博士生导师，博士。
收稿日期：2020-02-05 发表日期：2020-04-02 本文责任编辑：刘远颖



知识管理论坛
2020 年 第 2 期（总第 26 期）

对科技资源进行规范化，有效支持科技资源的集成、评价和共享。科技资源包括知识、数据、产品、人才、软件、硬件等不同类型。在这方面已经有不少的研究与应用，但还存在一些不足和进一步的需求：

（1）现有的研究主要集中在对不同类型的科技资源进行各自的描述，但缺乏对不同类型的科技资源进行统一描述，这对不同类型的科技资源的统一搜索和集成不利。例如，对知识图谱的研究较多 [1]，而科技资源图谱包括数据、产品、人才、软件、硬件等的"图谱"，这种研究还是比较缺乏。在中国知网中利用"科技资源图谱"作为主题词搜索，搜索到的结果为 0 条，而利用"知识图谱"作为主题词搜索，搜索到 10 542 条结果。

（2）人们已经对一些科技资源的分类模型、元数据模型、本体模型、知识元模型、知识图谱等进行了分别研究，并且已经有一些国家标准。王志强、杨青海等认为：科技资源开放共享过程中产生了数量庞大、种类繁杂的标准规范，这些标准规范对推动科技资源建设发挥了重要作用，但是也存在着一些问题，如缺乏全局性顶层设计，没有形成统一的标准化建设体系框架 [2]。需要进一步对这些模型进行集成统一研究，并用于科技资源的描述。

（3）在现有的研究中，对知识资源的描述已经有比较系统的方法，需要将这些方法扩展到其它类型的科技资源。

笔者将对这些问题进行研究，这有助于科技资源的描述方法的规范化、以及解决科技资源共享难的问题。

2 科技资源描述模型的结构框架

科技资源描述是科技资源的一种"画像"，是对科技资源的分类，是对科技资源的有序化，能够帮助用户快速搜索到所需要的科技资源和了解科技资源的主要内容，促进科技资源相互之间的快速集成，解决数据格式不一致和同一概念描述不一致的问题。图 1 为科技资源描述模型的结构框架，图 2 为科技资源描述模型间的关系。

科技资源描述模型中的各子模型定义如下：科技资源分类模型——描述科技资源的分类信息，以便找到所需要的科技资源；科技资源元数据模型——描述科技资源的主要数据格式，以便科技资源的快速集成；科技资源本体模型——对科技资源的规范性描述，以便准确、全面地找到所需要的科技资源；科技资源知识元——对科技资源内容进行简要描述，以便快速了解科技资源的主要内容；科技资源图谱模型——简要描述科技资源概念间的关系，以便科技资源的搜索和推理。

在科技创新和其他科技工作中，往往需要多种不同类型的科技资源的集成使用，如某研究任务，需要能胜任的研发人员（从人才资源中选择）、与研发任务相关的产品资源（参考相似产品，提高研发效率）、知识资源（如产品原理，产品可制造性、可装配性、可维护性等知识）、数据资源（如相似产品的历史使用数据、维护数据等）、软件资源（帮助研发的计算机辅助软件）、硬件资源（如实验设备、测试仪器等）。

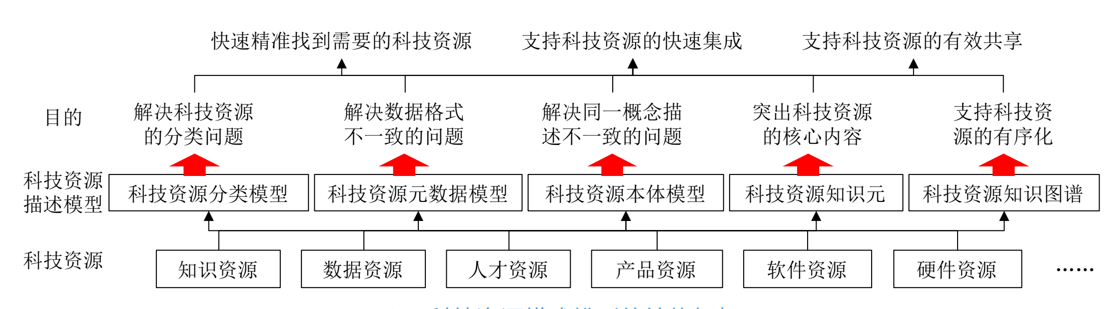


图 1 科技资源描述模型的结构框架

知识管理论坛，2020（2）:69-81
DOI：10.13266/j.issn.2095-5472.2020.007

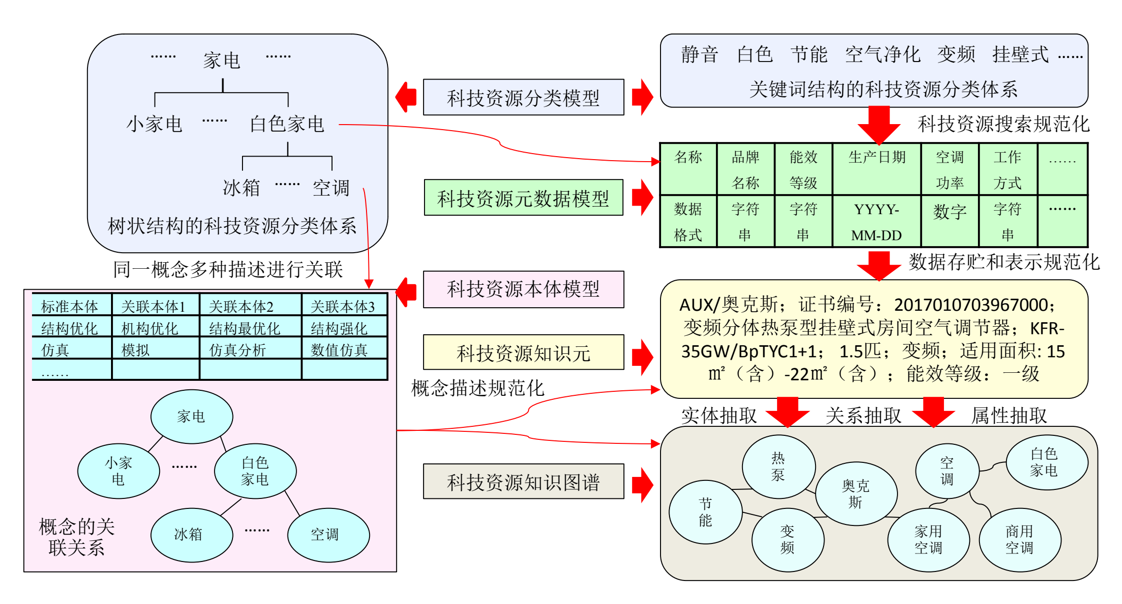


图 2 科技资源描述模型间的关系

3 科技资源分类模型及建立方法

3.1 科技资源分类模型的定义

科技资源分类模型是把具有某种属性或特征的科技资源信息归并起来，通过其属性或特征来区别不同类别的科技资源信息 [3]。根据不同的科技资源及需求，科技资源分类模型可以分为以下两种：

（1）树状结构的科技资源分类模型。这是采用数字或字母的形式，按照分类编码的一般原则与方法，对科技资源进行统一分类和编码，具有层次性和系统性的特点，可以确定任一科技资源在科技资源体系中的位置与相互关系。树状结构的科技资源分类模型又被称为科技资源分类编码体系 [4]、科技资源分类目录、科技资源标识体系等。与科技资源分类相近的分类编码体系有制造业信息化服务平台服务资源分类编码 [5]、网络化制造环境下的制造资源分类编码 [6]、企业信息分类编码 [7] 等。具体的科技资源的分类编码标准已经有工艺分类编码 [8]、零件分类编码 [9] 等。

树状结构的科技资源分类模型首先按照科技资源的性质不同进行基本分类。表 1 介绍了科技资源现有的一些分类理论，体现了科技资源分类的多样性。

表 1 科技资源现有的一些分类

分类理论	科技资源要素的主要内容
二要素论	科技信息资源、科技实物资源[10]；科技基础条件资源、技术创新资源等[11]
三要素论	科技实物资源、科技信息资源、科技服务资源[2]
四要素论	科技人力资源、财力资源、物力资源以及数字化时代的信息资源[12]； 人力资源、物力资源、财力资源、技术资源[13]
五要素论	科技人力资源、科技财力资源、科技装备资源、科技信息资源、科技政策与管理资源[14]； 人力、物力、财力、组织、管理、信息等资源[15]
七要素论	基础性核心科技资源（包括科技人力资源、科技财力资源、科技物力资源、科技信息资源）、整体功能性科技资源（包括科技市场资源、科技制度资源和科技文化资源）[16]
八要素论	大型科学仪器设备、重大科技基础设施、研究试验基地、自然科技资源、科学数据、科技图书文献、科技成果、科普资源等[17]



知识管理论坛
2020 年 第 2 期(总第 26 期)

国家标准《GB/T 32843-2016 科技资源标识》给出了科技资源标识方法，这是一种树状结构的科技资源分类模型，如图 3 所示：

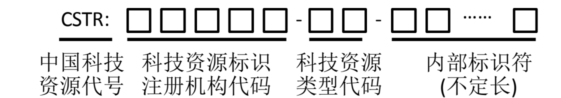


图 3 科技资源标识符结构（GB/T 32843–2016 科技资源标识）

其中：中国科技资源代号为 CSTR；科技资源标识注册机构代码为 5 位码；科技资源类型代码为 2 位码；内部标识符不定长，由科技资源标识注册机构分配，确保在同一科技资源标识注册机构注册的每个科技资源的内部标识符的唯一性。

该方法的缺点是：对于同一科技资源（如某科技文献），不同科技资源标识注册机构给出的科技资源标识符是不同的。但如果不考虑科技资源标识注册机构，面对如此众多的科技资源，要建立统一的科技资源标识符是很难的。

（2）关键词结构的科技资源分类模型。采用关键词或者标签 (tag) 等方式进行科技资源的属性或特征的表征和描述。这里的关键词或者标签往往是大众编制，所以又称大众分类法。这类分类体系比较适合互联网中的资源分享[18]。表 2 为两种科技资源分类模型的比较：

表 2 两种科技资源分类模型的比较

类别	树结构的科技资源分类模型	关键词结构的科技资源分类模型
主观性的影响	很强；难以表达成唯一的分类体系	较小；通过大数据的分析，减少个人主观性的影响，反映大众的选择
灵活性	较弱；修改难	很强；维护难
结构性	很强；树状结构本身就体现了分类结构的严谨性	较弱；关键词本身没有反映其相互关系，关键词之间的结构需要进行进一步的大数据分析才能得到
编制的复杂性	编制复杂要兼顾各种科技资源分类的需要，包括知识、数据、人才、产品、软件、硬件等科技资源。	编制简单
一致性	较强；由专家讨论统一确定	较弱；人们可能采用不同的术语描述同一概念
持续性	资源描述具有较长的持续性，可以有效保证其在时间历程上的一致性	有时关键词的描述随时间有较大变化，使过去的资源的搜索变得困难

3.2 科技资源分类模型的需求

科技资源分类模型的需求主要包括如下几个方面：

（1）有助于科技资源的统一有效组织管理和共享服务；

（2）通过建立科技资源的分级标准，支持科技资源的开放和共享；

（3）具有规范化和标准化的特性，支持科技资源的供需匹配；

（4）可以快速定位到所需要的科技资源，支持科技资源共享。

3.3 科技资源分类模型的建立方法

（1）树状结构的科技资源分类模型的建立方法。本文主要关注企业、平台的科技资源分类模型。因为国际、国家的科技资源分类模型比较宽泛，难以满足具体企业、平台的具体需求。

本文参考《中国图书馆分类法》《GB/T 32843-2016 科技资源标识》、国际专利分类体系 (IPC) 等分类体系，在此基础上进行扩展建立企业或行业平台的科技资源分类模型。《中国图书馆分类法》（简称《中图法》）是当今国内图书馆使用最广泛的分类法体系。目前国际上主要的专利分类体系有国际专利分类体系 (IPC)、日本专利分类体系 (FI/F-term)、美国专利分类体系 (USPC)、欧洲专利分类体系 (ECLA/ICO) 以及联合专利分类（CPC）等。在知识资源分类方面可以参考《GB/T 23703.7-2014 知识管理 第 7 部分：知识分类通用要求》。

知识管理论坛，2020（2）:69-81
DOI：10.13266/j.issn.2095-5472.2020.007

细分类别的科技资源可以参考一些现有标准，2020 年 1 月 29 日在国家标准信息查询平台（http://www.gov.cn/fuwu/bzxxcx/bzh.htm）输入"分类"搜索到国家标准 620 个、行业标准 704 个、地方标准 134 个。其中不少具有参考价值。

科技资源的树结构分类体系由本领域专家编制，将科技资源归入对应的子类，检索时可按树状结构一层一层地找到所需要的科技资源。科技资源的树结构分类体系的建立应遵循科学性、系统性、可延性和兼容性的原则，要尽可能请领域专家参与。

科技资源的内容和概念随时间不断变化，科技资源分类模型需要与时俱进，不断维护，或者在编码搜索系统中建立对应表，实现在不同时期的科技资源分类模型的统一搜索，这样可以解决传统的科技资源分类模型修改难、灵活性差等问题。例如，通过构建基于互联网的科技资源分类模型建立、维护和应用平台，来提高传统科技资源分类模型的灵活性和易维护性。

（2）关键词结构的科技资源分类模型的建立方法。主要采用大众分类法，即关键词或标签是由大众自己选择。①关键词的定义：出现在文献的标题、摘要以及正文中，能够表达文献主题内容、可作为检索入口的未经过规范化的自然语言词汇 [19]。②标签的定义：不依赖于固定分类，通过用户针对内容添加的简短描述，以方便搜索 [20]。

关键词结构的科技资源分类模型最大的问题是随意性较大、规范性较弱，这显著增加了搜索或匹配科技资源的难度。但在互联网环境中，随着关键词或标签用户数的增加，这种随意性将会显著减少，因为如果科技资源发布者所采用的关键词或标签太随意、不规范，就会使其发布的科技资源难以被人搜索和利用，达不到其发布科技资源的目的；同样，如果科技资源搜索者所采用的关键词或标签不规范，就会使其难以搜索到想要的科技资源。最终对于同一科技资源，大家就会趋向于采用同样的关键词或标签。这是一种自组织优化的模式，互联网平台要为促进关键词或标签的自组织优化提供良好的环境。例如，当用户输入关键词或标签时，平台提示该关键词或标签是否是常用的，并根据科技资源的特点智能推荐常用的关键词或标签。

关键词或标签可以采用本体模型进行优化，提高基于关键词或标签的科技资源的搜准率和搜全率，具体见第 5 节。

4 科技资源元数据模型及建立方法

4.1 科技资源元数据模型的定义

科技资源元数据规范了科技资源描述空间的维度，是描述数据的数据（data about data），用于描述科技资源（包括实物资源和信息资源）的内容、覆盖范围、质量、管理方式、数据的所有者以及提供方式等有关信息的数据 [28]。关于元数据有不同的定义：

（1）关于数据的数据 [21]。

（2）定义和描述其他数据的数据 [22]。

（3）关于数据或数据元素的数据（可能包括其数据描述），以及关于数据拥有权、存取路径、访问权和数据易变性的数据 [23]。

（4）描述数据及其环境的数据 [24]。

（5）描述物联网数据及其相关信息的数据 [25]。

（6）关于数据的数据，主要是描述数据属性（property）的信息 [26]。

（7）描述科技报告的一种结构化数据，用于实现检索、管理、使用、保存等功能 [27]。

这种元数据定义的多义性说明了建立统一的科技资源元数据的难度。

科技资源的元数据包括：科技资源名称、类型、发布者、发布时间、存放地点、关键词等 [28]。对不同的科技资源（如知识、数据、人才、产品、软件、硬件等）有相应的元数据模型，有些已经有标准，需要考虑尽可能采用。

4.2 科技资源元数据模型的需求

不同的人对科技资源描述空间的维度往往

有不同的定义，这就导致了科技资源集成难和搜索难。科技资源元数据通过对科技资源对象进行统一规范描述，有助于对科技资源的组织、集成、检索、发现和管理 [30]。

4.3 科技资源元数据模型的建立方法

（1）参考已有的科技资源元数据模型，调查搜集尽可能多的科技资源元数据，建立科技资源元数据参考模型库。2020 年 1 月 29 日在国家标准信息查询平台（http://www.gov.cn/fuwu/bzxxcx/bzh.htm）输入“元数据”搜索到国家标准 66 个、行业标准 53 个、地方标准 24 个。例如，目前已经有《GB/T 36478.3-2019 物联网 信息交换和共享 第 3 部分：元数据》《GB/T 38154-2019 重要产品追溯 核心元数据》《GB/T 37282-2019 产品标签内容核心元数据》《GB/T 37600-2018 全国主要产品分类 产品类别核心元数据》《GB/T 35430-2017 信息与文献 期刊描述型元数据元素集》《GB/T 35397-2017 科技人才元数据元素集》《GB/T 30535-2014 科技报告元数据规范》《GB/T 30523-2014 科技平台 资源核心元数据》《GB/T 30522-2014 科技平台 元数据标准化基本原则与方法》《GB/T 30522-2014 科技平台 元数据标准化基本原则与方法》《GB/T 26499.3-2011 机械 科学数据 第 3 部分：元数据》《GB/T 25100-2010 信息与文献 都柏林核心元数据元素集》《GB/T 24662-2009 电子商务 产品核心元数据》《GB/T 18391-2009 信息技术 元数据注册系统 (MDR)》《GB/T 22373-2008 标准文献元数据》《GB/T 22373-2008 标准文献元数据》等。

（2）从科技资源元数据参考模型库中，根据需要选择合适的科技资源元数据。如果元数据数量太多，使用不便；元数据数量太少，则描述不完整。需要进行元数据的相关性分析，去掉相关性较大的两个元数据中的一个；需要进行元数据的重要性评价，把对科技资源描述价值相对较小的元数据去掉；元数据的数量最终要考虑科技资源描述的完整性、特征可识别性、可分类性等；元数据的数量还与科技资源的其他具体描述需求有关；元数据选择与元数据建立和管理的信息化水平有关，当信息化较高时，元数据的数量可以多些。

（3）科技资源元数据类型可以由专家协商确定，也可以通过大数据分析得到，或者由专家协商和大数据分析共同得到。

（4）协同建立科技资源元数据模型的标准。该标准涉及面广、用户多，因此可以采用维基（Wiki）模式，组织广大用户参与，协同提出和修改科技资源元数据模型的标准。

科技资源核心元数据的定义是：描述科技资源最基本信息的元数据最小集合（修改自 GB/T 30523-2014 科技平台 资源核心元数据），包括：科技资源中文名称和英文名称、科技资源发布者、科技资源发布时间（最近提交日期）、科技资源存放地点（信息链接地址）、科技资源知识元、科技资源关键词（或标签）、科技资源标识编码、科技资源标准本体和关联本体。

元数据建立方法可以参考《GB/T 30522-2014 科技平台 元数据标准化基本原则与方法》《GB/T 26499.3-2011 机械 科学数据 第 3 部分：元数据》。

5 科技资源本体模型及建立方法

5.1 科技资源本体模型的定义

本体没有统一的定义，以下给出一些不同领域的国家标准对本体的定义：

（1）在大数据语境下，它是一些约束后续各种不同层次逻辑模型的语义模型 [19]。

（2）计算机科学领域的一种模型，用于描述用一套对象类型（概念或者说类）、属性以及关系类型所构成的世界 [31]。

（3）被表述为一系列相互关联的概念与定义，这种表述类似于叙词表中的术语。但是，本体不是术语标准 [32]。

（4）使用计算机能够处理的语言对论域的描述 [33]。

（5）在文化遗产信息资源领域，基于本体的模型用于将异构、分散的文化遗产信息源进

知识管理论坛，2020（2）:69-81
DOI：10.13266/j.issn.2095-5472.2020.007

行集成、交换，有助于形成通用的、规范的本体模型，给领域专家对信息的编制和关联检索提供指南[34,36]。

许多领域需要通过本体构建，实现相关业务的标准化工作；同时，本体也是基于 Web 应用的互操作问题的关键。因此业界正在陆续制定相关的本体标准。

本文对科技资源本体定义是：科技资源本体模型规范了同一科技资源的名称术语及不同名称术语间的关系。

5.2 科技资源本体模型的需求

不同的人对科技资源往往有不同的名称术语及名称术语的关系，这就导致了科技资源集成难和搜索难。面对庞大的科技资源和名称，主要存在以下问题：

（1）有时同一科技资源有多种名称术语，一种名称术语描述多种不同的概念，这对科技资源的集成带来诸多不便。一方面需要通过标准化、规范化的方法解决这些问题，例如，采用数据字典[35]-[36]等方式；另一方面可以通过本体方法，建立标准本体和关联本体。标准本体对应描述某一概念的标准术语，关联本体对应描述这一概念的其他术语。在信息搜索时标准本体与关联本体一起用于搜索。科技资源本体模型通过对科技资源对象进行统一规范的描述，有助于对科技资源的组织、集成、检索、发现和管理。

（2）有时同一科技资源有多种概念结构，这对科技资源的集成也带来诸多不便。本体有助于解决同一概念的名称多样化问题和概念结构混乱带来的问题。名称多样化问题会进一步导致科技资源共享和利用中出现如下问题：①搜索到的科技资源信息不完整；②搜索到的科技资源信息不准确；③科技资源信息集成难。概念结构混乱会带来科技资源分类混乱、资源集成难和搜索难的问题。而科技资源本体模型，有利于实现科技资源的共享、集成、服务，例如：

- 知识资源共享：满足企业知识资源库统一检索、企业知识图谱建立、技术路线图共建、技术进化图共建、知识推送等需求，提高知识资源的有序化程度。
- 人才资源共享：支持对人才资源的统一描述和搜索。
- 软件资源共享：支持对软件资源的统一描述和搜索；满足不同阶段和不同单位开发的不同的软件系统之间集成的需要，主要是不同数据库中的字段名的映射、不同数据结构的映射等的需求。
- 人工智能系统：支持知识间逻辑关系的建立、推理机的实现，满足人工专家系统、智能辅助决策系统等的建立等。

5.3 科技资源本体模型的建立方法

科技资源本体模型包括标准本体和关联本体，其概念如图 4 所示：

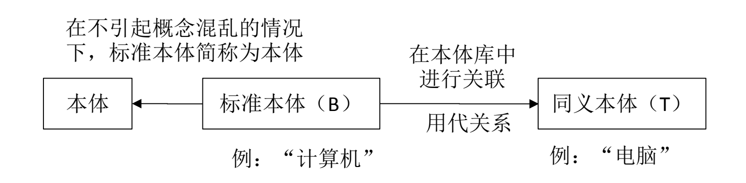


图 4 本体、标准本体和同义本体的关系

科技资源本体模型的建立方法主要包括：

（1）了解企业的业务组织及工作内容，确定企业所需要共享的科技资源范围；确定科技资源本体的需求。

（2）初选科技资源本体，试用和选择科技资源本体，包括标准本体和关联本体。

（3）依靠广大科技人员协同建立科技资源本体，并通过对大家使用科技资源的行为的跟踪、统计和分析，不断优化科技资源本体[37]

标准本体与术语概念类似，可以参考《GB/

T 10112-2019 术语工作 原则与方法》《GB/T 13725-2019 建立术语数据库的一般原则与方法》。

2020 年 1 月 29 日在国家标准信息查询平台（http://www.gov.cn/fuwu/bzxxcx/bzh.htm）输入“术语”搜索到相关的国家标准 1 172 个，行业标准 840、地方标准 932。绝大多数是各种产品、技术的术语标准。但其给出的术语数量还比较少，在本体建设中，需要适当扩充。

p- 关联本体则是术语标准中所缺乏的，其有助于提高科技资源的搜准率和搜全率，需要花费较多精力从术语的同义词、近义词中寻找。

6 科技资源知识元及建立方法

6.1 科技资源知识元的定义

科技资源知识元 [19] 是从科技资源中进一步提炼而成的科技资源中的最核心和最精炼的知识，往往是以摘要、简要介绍等方式展示。

已有标准将知识元定义为：在应用需求下，表达一个完整事物或概念的不必再分的独立的知识单元 [19]。

科技资源知识元的内容主要是：

（1）目的 / 意义：简要说明科技资源的需求、干什么用（为什么，Why）。例如，某科学仪器检测的目的是什么。

（2）方法 / 过程：简要说明科技资源的建立和应用方法（怎么用，How）。例如，某科学仪器的检测原理及检测精度。

（3）结果 / 结论：简要说明科技资源的内容和应用结果（是什么，What）。例如，某科学仪器的具体检测内容，检测后可以得到什么结果。

6.2 科技资源知识元的需求

知识元首先是从文献领域发展起来的。早在 20 世纪 70 年代后期，美国专家指出：文献数量膨胀之后，知识的控制单位将从文献深化到文献中的数据、公式、事实、结论等最小的独立的“知识元”，知识元可以被称为是文献管理的最小单位。知识元不仅可以用于情报管理中的文献处理，而且，知识元也可以表示其他种类知识载体，如专利等，将其中所涉及的概念、论据、论证以及创新点等知识核心以知识元的方式呈现，以此作为知识管理、知识评价以及知识发现的最小单元 [37,43]。

科技资源知识元可以让用户快速地了解有关科技资源的主要特点和内容，仅仅依靠关键词等是难以了解科技资源的大致面貌。科技资源知识元可以支持科技资源知识图谱的快速构建，支持科技资源知识元之间的快速集成。

6.3 科技资源知识元的建立方法

科技资源知识元的内容主要是：

（1）简要说明科技资源的需求（为什么，Why）。例如，某科学仪器检测的目的是什么。

（2）简要说明科技资源的内容（是什么，What）。例如，某科学仪器的具体检测内容。

（3）简要说明科技资源的应用方法（怎么用，How）。例如，某科学仪器的检测原理及检测精度。

为了提高科技资源的搜索和利用效率，需要按照科技资源元数据模型，采用标准本体描述科技资源。

7 科技资源图谱及建立方法

7.1 科技资源图谱的定义

知识图谱实质上是一种构建实体间关系的语义网络，它可以形式化地描述客观世界中的事物及其相互关系。如今，知识图谱被用来指代各种大规模的知识库。2012 年，谷歌率先提出了知识图谱的概念，旨在增强搜索引擎的理解能力，提高搜索质量和用户体验 [38]，此后，知识图谱的研究方向受到了广泛的关注。知识图谱以其强大的开放性、互联性和语义处理能力为互联网中的知识互联奠定了基础。

三元组是知识图谱的一种通用的表示方式，即：$G=(E, R, S)$。其中，$E=\{e_1, e_2, \cdots e_{|E|}\}$ 表示知识库中的实体集合，共包含 $|E|$ 种不同的实体；$R=\{r_1, r_2, \cdots r_{|R|}\}$ 表示知识库中的关系集合，共包含 $|R|$ 种不同的关系；$S \subseteq E \times R \times E$ 代表知识库

知识管理论坛，2020（2）:69–81
DOI：10.13266/j.issn.2095–5472.2020.007

中的三元组集合。三元组的基本形式主要包括实体（Entity）-关系（Relationship）-实体（Entity）和（实体-属性-属性值）等。每个实体（概念的外延）可用一个全局唯一确定的ID来标识，每个属性-属性值对可用来刻画实体的内在特性，而关系可用来连接两个实体，刻画它们之间的关联。

科技资源图谱的概念是在知识图谱基础上发展起来的，用于显示科技资源发展进程与结构关系的一系列各种不同的图形模型，采用可视化技术描述知识资源及其载体，挖掘、分析、构建、绘制和显示知识及它们之间的相互联系，是对科技资源的全方位关联关系的描述 [39]。

上海人工智能公共研发资源图谱已经收录人工智能及相关领域的专家人才信息超过10万条、学科词库超过30万条、科技文献超过1亿篇、科技机构超过10万家、科技企业超过1万家，为全球科研从业人员带来全新的知识搜索服务体验以及基于深度数据分析产生的科研趋势可视化分析，帮助科研人员更快、更丰富、更精准地寻找专业科研资源并发现科研热点和未来方向，有效增加科研人员的工作效能和成果 [40]；SciKG[41] 是清华大学计算机科学领域研发的知识图谱，图谱由概念、专家、论文等元素构成，专家和论文都有相应规定的一些属性，将专家和论文关联起来，用于帮助研究人员更好地搜索计算机领域的专家和论文等；gstore[42] 是北京大学建立的图数据库，结合RDF存储和SPARQL查询，支持海量的三元组知识图谱数据管理，并被应用于全球微生物中心知识图谱构建和方正智汇对出版资源的统一管理中。科技资源图谱的内容和需求如图5所示：

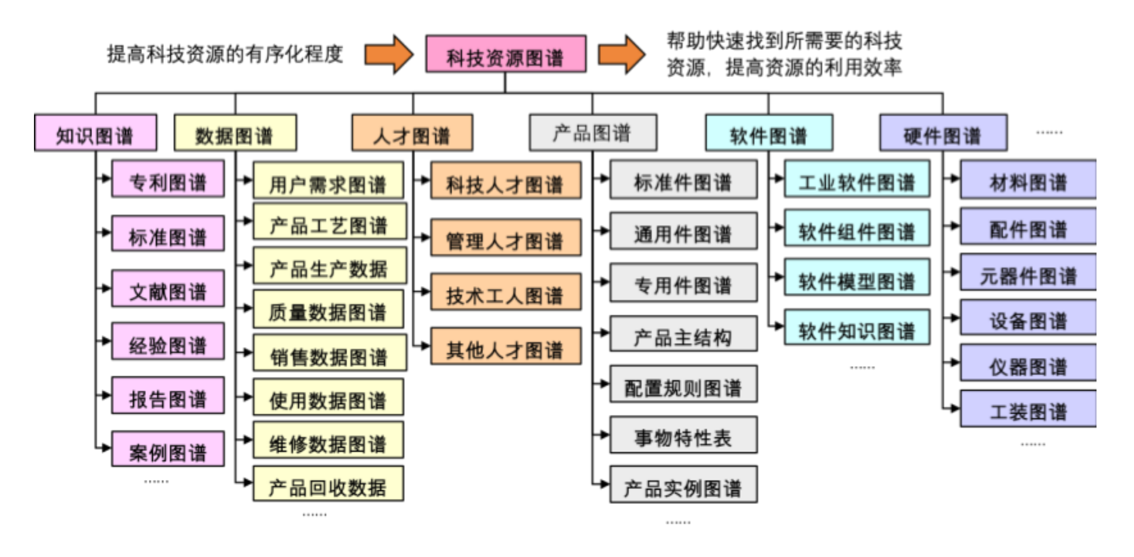


图5 科技资源图谱的内容和需求

（1）知识图谱。以知识分类体系或关键词为核心，建立知识之间的各种关系，如关联关系、层次关系、衍生关系、相似关系等，集聚知识的属性。

（2）数据图谱。说明数据之间的关系，例如，面向某机床的加工质量的原因分析的数据，包括：机床振动数据、机床热变形数据、刀具加工声发射数据、刀具磨损视觉监控数据、工件加工表面质量数据等，由数据图谱集成，目的是便于数据的管理和利用。数据图谱还关联获取这些数据的人、传感器、软件等，关联相应的机床、刀具、工件等参数。其目的是使这些数据能够被大家共享重用，提高数据的价值。

（3）人才图谱。以知识分类体系或关键词为核心，建立人才之间的各种关系，如师生关系、合作伙伴关系、竞争对手关系、专业相似关系、

专业互补关系等，并集聚人才的各种成果。

（4）产品图谱。以产品分类体系或关键词为核心，建立产品之间的各种关系，如层次关系、相似关系、成套关系、变型关系、配置关系等，并集聚产品的各种信息。

（5）软件图谱。建立软件之间的各种关系，如可组合关系、可变型关系、可置换关系等，并集聚软件的各种相关描述信息。

（6）硬件图谱。建立硬件之间的各种关系，如层次关系、相似关系、成套关系等，并集聚硬件的各种相关描述信息。硬件种类很多，差别很大，所以首先需要对硬件进行分类。

7.2 科技资源图谱的需求

科技资源之间具有一定的关联性，可以采用科技资源图谱对其进行描述。利用科技资源图谱可以帮助快速搜索到系统化的科技资源，提高科技资源的利用效率。例如，数据之间的关系通过数据图谱可以完整获得。

通过科技资源图谱可以有序化集成和全方位描述科技资源，方便大家共享。

7.3 科技资源图谱的建立方法

科技资源描述的难点是科技资源描述的工作量很大，并会因人而异，需要采用透明公平的方法激励大家参与科技资源描述，需要采用大数据和群体智能的方法提高科技资源描述的自动化水平和准确性。

不同的专家由于自己所擅长的细分领域的不同、科研水平和素养的不同，所以在科技资源图谱建立中需要给予不同的权重。

知识图谱的构建模式分为自顶向下（top-botton）和自底向上（botton-top）两种。自顶向下指首先定义本体库和数据模式，再向知识库中添加一系列事实，即先模式层后数据层。自底向上指先提取文本分析数据，再由数据驱动，设计知识库的模式层，即先数据层后模式层。一般的知识图谱是自底向上构建的，比如谷歌的 Knowledge Vault 知识库。然而，对于垂直领域知识图谱，在处理复杂和不稳定的业务需求时，需要特定于行业的专业知识和高质量的数据，则更倾向于采用自顶向下的方法。

8 结语

创新是我国的发展战略，创新尤其是协同创新需要科技资源共享，而科技资源共享的前提是要有一套行之有效的规范、合理、科学的科技资源描述模型和建立方法。

笔者提出一套科技资源描述模型的结构框架，其特点是对科技资源从不同角度进行规范化，形成一个整体、系统的描述，有效支持科技资源的集成、评价和共享，包括：

（1）科技资源分类模型。主要是树结构和关键词两种分类模型，它们各有优缺点，可以互补。建议以树结构分类模型为主，关键词分类模型为辅，以便适合大范围、跨专业的科技资源分类。

（2）科技资源元数据模型。从不同种类的科技资源集聚和共享的需求出发，提出统一的科技资源元数据模型。

（3）科技资源本体模型。面对庞大的科技资源概念和名称，存在的问题是：有时同一概念有多种名称，这对科技资源的集成带来诸多不便。一方面需要通过标准化、规范化的方法解决这些问题，例如，采用术语标准、数据字典等方式；另一方面可以通过本体方法，建立标准本体和关联本体。标准本体对应描述某一概念的标准术语，关联本体对应描述这一概念的其他术语。在信息搜索时标准本体与关联本体一起用于搜索。

（4）科技资源知识元。将科技资源的主要内容简要描述出来，方便使用，支持科技资源图谱的建立。

（5）科技资源图谱。将知识、数据、产品、人才、软件、硬件等科技资源采用图谱的方式进行关联和可视化，使科技资源之间的关系清晰化，使围绕某一任务的科技资源集聚为一个整体，方便科技资源的搜索和利用。

上述模型对科技资源描述提供了一个比较规范、简要和完整的整体解决方案，有助于提

知识管理论坛，2020（2）:69–81
DOI：10.13266/j.issn.2095–5472.2020.007

高科技资源的集成、评价和分享能力。

笔者还提出一套科技资源描述模型的建立方法，其特点是利用新一代信息技术，依靠大众共建模型，依靠科技资源描述过程的大数据智能建立和优化模型。科技资源描述模型大多是科技资源共享中的基础标准，这些标准很多，并且经常变化，需要通过开放、分布、并行、协同、智能的方法共建。

开放的方法是指这些标准建设开放给感兴趣的企业，大家一起参与。分布的方法是指这些标准建设者是平等的，谁贡献大，谁就是标准起草者；标准起草者按照贡献大小排名。并行的方法是指这些标准的建设与相关系统的建立和开发并行的，不是等到方法和技术已经很成熟了，再建标准。协同的方法是指这些标准建设者相互协同，资源共享，提高标准的水平，缩短标准建设周期。智能的方法是指这些标准建设过程利用大数据分析方法，简化标准建设的工作量；智能地监控标准建设工作，每个人的贡献透明，排名公平。

参考文献：

[1] 黄恒琪 , 于娟 , 廖晓等 . 知识图谱研究综述 [J]. 计算机系统应用 , 2019, 28(6): 1-12.

[2] 王志强 , 杨青海 . 科技资源开放共享标准体系研究 [J]. 中国科技资源导刊 , 2016, 48(4): 19-23.

[3] 董明涛 , 孙研 , 王斌 . 科技资源及其分类体系研究 [J]. 合作经济与科技 , 2014(10): 28-30.

[4] 国家质量监督检验检疫总局 , 国家标准化管理委员会 . 企业信息分类编码导则第 1 部分 : 原则与方法 :GB/T 20529.1-2006[S]. 北京 : 中国标准出版社 , 2007.

[5] 国家质量监督检验检疫总局 , 国家标准化管理委员会 . 制造业信息化服务平台服务资源分类规范 :GB/T 34045-2017[S]. 北京 : 中国标准出版社 , 2018.

[6] 国家质量监督检验检疫总局 , 国家标准化管理委员会 . 网络化制造环境下的制造资源分类 :GB/T 25111-2010[S]. 北京 : 中国标准出版社 , 2010.

[7] 国家质量监督检验检疫总局 , 国家标准化管理委员会 . 企业信息分类编码导则第 2 部分 : 分类编码体系 : GB/T 20529.2-2010[S]. 北京 : 中国标准出版社 , 2011.

[8] 国家质量监督检验检疫总局 , 国家标准化管理委员会 . 面向装备制造业产品全生命周期工艺知识第 2 部分 : 通用制造工艺分类编码规范 : GB/T 22124.2-2010[S]. 北京 : 中国标准出版社 , 2011.

[9] 陕西省市场监督管理局 . 基于成组技术的零件分类编码要求 :DB61/T 1224-2018[S]. 北京 : 中国标准出版社 , 2018.

[10] 涂勇 , 龚雪娟 , 赵辉 . 科技资源管理标准体系的研究 [J]. 中国科技资源导刊 , 2012(6): 41-44.

[11] 国家质量监督检验检疫总局 , 国家标准化管理委员会 . 科技资源标识 :GB/T 32843-2016[S]. 北京 : 中国标准出版社 , 2016.

[12] 孙凯 . 科技资源共享可行性分析及对策建议 [J]. 西北大学学报 (哲学社会科学版), 2005, 35(3): 109-112.

[13] 王雪 . 区域科技共享平台服务模式与运行机制研究 [D]. 哈尔滨 : 哈尔滨理工大学 , 2015.

[14] 范菲雅 , 倪炎榕 , 袁晓舟 , 等 . 网络化制造环境下基于语义 Web 的应用服务资源模型 [J]. 计算机集成制造系统 , 2009(8): 53-59.

[15] 国家质量监督检验检疫总局 , 国家标准化管理委员会 . 科技平台资源核心元数据 :GB/T 30523-2014[S]. 北京 : 中国标准出版社 , 2015.

[16] 刘玲利 . 科技资源要素的内涵——分类及特征研究 [J]. 情报杂志 .2008(8): 125-126.

[17] 国家质量监督检验检疫总局 , 国家标准化管理委员会 . 科技资源标识 :GB/T 32843-2016[S]. 北京 : 中国标准出版社 , 2017.

[18] 顾复 , 陈芨熙 . 一种基于标签的产品和零部件网页的自组织分类编码方法 [J]. 成组技术与生产现代化 , 2007, 24(2): 57-60.

[19] 国家市场监督管理总局 , 国家标准化管理委员会 . 新闻出版知识服务知识资源建设与服务基础术语 :GB/T 38377-2019[S]. 北京 : 中国标准出版社 , 2020.

[20] 国家质量监督检验检疫总局 , 国家标准化管理委员会 . 信息与文献期刊描述型元数据元素集 :GB/T 35430-2017[S]. 北京 : 中国标准出版社 , 2018.

[21] 国家质量监督检验检疫总局 , 国家标准化管理委员会 . 信息技术元数据注册系统 (MDR): GB/T 18391-2009 [S]. 北京 : 中国标准出版社 , 2010.

[22] 国家质量监督检验检疫总局 , 国家标准化管理委员会 . 信息技术词汇第 17 部分 : 数据库 :GB/T 5271.17-2010[S]. 北京 : 中国标准出版社 , 2011.

[23] 国家质量监督检验检疫总局 , 国家标准化管理委员会 . 物联网术语 :GB/T 33745-2017[S]. 北京 : 中国标准

出版社, 2018.

[24] 国家质量监督检验检疫总局, 国家标准化管理委员会. 物联网信息交换和共享第 3 部分: 元数据 :GB/T 36478.3-2019[S]. 北京: 中国标准出版社, 2020.

[25] 国家质量监督检验检疫总局, 国家标准化管理委员会. 科技人才元数据元素集 :GB/T 35397-2017[S]. 北京: 中国标准出版社, 2018.

[26] 国家质量监督检验检疫总局, 国家标准化管理委员会. 科技报告元数据规范 :GB/T 30535-2014[S]. 北京: 中国标准出版社, 2015.

[27] 刘春燕, 安小米. 基于生命周期的科技信息资源共享元数据研究 [J]. 情报理论与实践, 2018, 41(5): 39-43.

[28] 国家质量监督检验检疫总局, 国家标准化管理委员会. 科技平台资源核心元数据 :GB/T 30523-2014[S]. 北京: 中国标准出版社, 2015.

[29] 赵启阳, 张辉, 王志强. 科技资源元数据标准研究的现状分析与新的视角 [J]. 标准科学, 2019(3): 12-17

[30] 国家质量监督检验检疫总局, 国家标准化管理委员会信息技术大数据术语 :GB/T 35295-2017[S]. 北京: 中国标准出版社, 2018.

[31] 国家市场监督管理总局, 国家标准化管理委员会. 信息与文献文化遗产信息交换的参考本体 :GB/T 37965-2019[S]. 北京: 中国标准出版社, 2020.

[32] 国家市场监督管理总局, 国家标准化管理委员会. 信息技术互操作性元模型框架(MFI)第 3 部分: 本体注册元模型 :GB/T 32395-2015[S]. 北京: 中国标准出版社, 2016.

[33] 国家市场监督管理总局, 国家标准化管理委员会. 智能运输系统　数据字典要求 :GB/T 20606-2006[S]. 北京: 中国标准出版社, 2007.

[34] 国家市场监督管理总局, 国家标准化管理委员会. 新闻出版 知识服务 知识元描述 :GB/T 38381-2019[S]. 北京: 中国标准出版社, 2020.

[35] 国家市场监督管理总局, 国家标准化管理委员会. 基础地理信息要素数据字典　第 1 部分: 1 ∶ 500　1 ∶ 1000　1 ∶ 2000 比例尺 :GB/T20258.1-2019[S]. 北京: 中国标准出版社, 2020.

[36] 顾新建, 马步青, 代风. 基于大数据的知识共享方法研究 [J]. 知识管理论坛, 2016(1): 30-38.

[37] 毕经元. 基于 Web2.0 的知识元链接网络系统 [D]. 杭州: 浙江大学, 2010.

[38] AMITS. Introducing the knowledge graph[R]. America:Official Blog of Google, 2012.

[39] 杜鹏程, 吴婷, 王成城. 科技人力资源研究领域的知识图谱分析 [J]. 中国科技论坛, 2013(8): 83-89.

[40] 马亚宁. 上海人工智能公共研发资源图谱 [N]. 新民晚报, 2019-08-30(1).

[41] TANG J, ZHANG J, YAO L, et al. ArnetMiner: extraction and mining of academic social networks[C]// Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining.New York:Association for Computing Machinery, 2008: 990-998.

[42] ZOU L, ÖZSI M T, CHEN L, et al. gStore: a graph-based SPARQL query engine[J]. The VLDB journal, 2014, 23(4): 565–590.

[43] 毕经元, 顾新建, 吕艳, 等. 基于知识元链接的汽车零部件知识管理系统 [J]. 浙江大学学报(工学版), 2009, 43(12): 2208-2213.

作者贡献说明:

顾　复: 标准和其他文献的分析, 论文的写作;

刘杨圣彦: 论文部分内容的写作, 文献查阅;

顾新建: 提出论文的总架构, 修改论文。

知识管理论坛，2020（2）:69-81
DOI：10.13266/j.issn.2095-5472.2020.007

Description Method of Scientific and Technological Resources

Gu Fu Liu Yangshengyan Gu Xinjian
School of Mechanical Engineering, Institute of Industrial Engineering, Zhejiang University, Hangzhou 310027

Abstract: [Purpose/significance] Scientific and technological innovation is the key point of development, and scientific and technological resources sharing and collaborative innovation are indispensable. Sharing of scientific and technological resources is a systematic project, and first of all, in order to integrate, evaluate and share scientific and technological resources, we need to establish a description model of scientific and technological resources. **[Method/process]** The description model of science and technology resources defined in this paper includes classification model of scientific and technological resources, meta data model of science and technology resources, ontology model of science and technology resources, knowledge meta model of science and technology resources, science and technology resource graph model, etc. Scientific and technological resources include knowledge, data, products, talents, software, hardware and other resources. We proposed the method of establishing the model of science and technology resources description, as well as the characteristic and function of it. **[Result/conclusion]** The main contributions of this paper are as follows: standardizing the description model of science and technology resources is helpful for the integration and sharing of different types of science and technology resources; by the integration research of different types of description models of science and technology resources, we form the system architecture of the description model of science and technology resources, which provides an overall solution for the comprehensive and systematic description of science and technology resources, and helps us to share the science and technology resources; we put forward a method to build a description model of science and technology resources, based on crowd-sourcing theory and big data AI of description process of scientific and technological resources, and we establish and optimize the description model of science and technology resources.

Keywords: science and technology resource science and technology resource description model method of modeling science and technology resource sharing knowledge graph